

## 지능형 학회관리 시스템 설계 및 구현

조영기\*, 백성욱\*, 방기천\*\*

### 요약

본 논문에서는 다양한 연구 활동을 유도하기 위해 구축된 지능형 학회 관리 시스템의 설계 및 구현 사례를 다루었다. 본 시스템은 회원 정보와 학회활동 기록 등의 데이터를 기반으로 데이터마이닝을 수행하여 데이터마이닝 과정에서 나타난 여러 유용한 학회활동 규칙들을 생성함으로써 회원들과 효과적으로 상호작용 할 수 있도록 했다. 데이터마이닝은 여러 기법들 중 의사결정나무알고리즘(Decision Tree Algorithm)을 사용했다. 분석된 정보를 이용해 사용자 위주의 운영정책과 동적 인터페이스를 제공하기 위한 웹사이트 개인화 계획을 제시한다.

## Design and Implementation of Intelligent Society Member Management System

Yung-Ki Jo\*, Sung-Wook Baik\*, Kee-Chun Bang\*\*

### Abstract

This paper presents a design and implementation example of intelligent society member management system that is constructed to induce various research activity. Based on members data and society activity record, the system executed data mining. In the process of data mining, useful society activity rules was produced and in result members could effectively interact with the system. Decision Tree Algorithm was used in the process, which is one of the methods of data mining. We presents a plan for personalization website to provide user oriented administration policy and dynamic interface by using analyzed information of society activity rules produced.

Key words : 데이터마이닝, 의사결정나무알고리즘(Decision Tree Algorithm)

### 1. 서론

오늘날 디지털 정보기술의 발달로 정보관리와 활용에 대한 인식이 높아지면서 효과적인 정보관리와 정보활용 방안에 대한 연구가 활발해지고 있다. WWW의 규모와 사용의 지속적인 증가로 인해 인터넷을 통한 콘텐츠의 형식과 양, 그에 따른 웹 페이지들의 증가는 온라인 정보 서비스의 개발과 설계의 새로운 방법을 요구하고 있다. 많은 웹사이트들이 사용자와 보다 효율적인 상호작용을 위해 사용자들이 웹과 상호작용하는 동안 축적된 로그 정보를 활용해 웹사이트의 운영에 반영하려는 노력을 하고 있다. 초기에는 주로 로그 분석도구를 이용해 통계량을 통한 트래픽과 클릭에 따른 이동정보만을 분석했으나 이것만으로는 유용한 지식을 얻기 힘들었다. 로그 정보를 데이터 마이닝에 적용함으로써 여러 가지 상황에 대한 적절한 의사결정을 위한 의미있

는 고급 정보 혹은 지식을 발견하려는 연구가 활발하게 진행되고 있다. 특히 웹상에서는 사이트 간에 이동이 쉽고 사용자들의 유동성이 강함으로 사용자들의 요구사항과 특징들을 잘 이해해서 사용자에게 따른 개인화 방안을 강구하는 것은 매우 중요하다. 온라인에서 개인화를 활용하는 것은 실시간으로 개인화 기법을 적용하며 고객에게 새로운 것을 소개하고 고객의 잠재적 요구를 통해서 웹사이트의 내용을 변화시키는 것을 말한다. 이는 실시간으로 축적된 데이터를 데이터마이닝 기법을 이용해 데이터의 패턴을 분석함으로써 가능하다.[1]

데이터마이닝 기법을 이용한 지능형 개인화 엔진을 학회 관리 시스템에 활용하기 위해 사용자들의 성향을 파악하여 사용자들에게 필요한 콘텐츠와 정보를 제공한다면 적극적인 연구 활동을 유도할 수 있을 것이다.

본 논문에서는 지능형 학회 관리 시스템의 설계 및 구현 사례를 제시한다.

\* 제일저자(First Author) : 조영기

접수일 : 2004년 7 월 25 일, 완료일 : 2004년 8 월 20 일

\* 세종대학교 디지털콘텐츠학과 석사과정

\* 세종대학교 컴퓨터 공학부 디지털콘텐츠학과 교수

\*\* 남서울대학교 멀티미디어학과 교수

학회 관리 시스템을 통해 수집한 웹 로그 정보와 학회 데이터베이스를 통해 얻은 사용자의 인구 통계학적 정보들의 상관관계를 데이터마이닝 과정에 따라 분석하여 여러 패턴들을 발견했으며 분석된 결정규칙(Decision Rule)을 기반으로 사용자 위주의 운영정책과 동적 인터페이스를 제공하기 위한 웹사이트 개인화(Personalization) 계획을 제시한다.

## 2. 시스템 설계 및 구현

### 2.1 시스템의 개요

지능형 학회 관리 시스템은 사용자들이 인터넷을 통해서 손쉽게 각종 콘텐츠를 접할 수 있도록 구현되었다. 회원은 등급별로 나누어 정회원, 준회원, 학생회원, 임원 등으로 나누어 관리하며 각각 회원별로 제공되는 기능과 콘텐츠가 상이하다. 이처럼 각 회원별로 차별화된 서비스를 제공함으로써 콘텐츠의 열람자, 저작자, 편집자, 승인자들과 어플리케이션 사이의 작업과정을 자동화할 수 있도록 설계했다. 논문에 대한 정보는 XML 형태로 구조화시켰으며 각 논문의 정보를 파싱(parsing)을 통해 제공하여 콘텐츠가 제공되는 메뉴 구조, 논문이 표현되는 디자인 요소 그리고 실제 제공되는 논문내용을 분리해서 관리하기 때문에 동일한 논문을 상이한 View를 통해 출판할 수 있도록 디자인 요소와 콘텐츠를 분리했다.

구현 환경은 Linux 7.2 운영체제 상에서 PHP와 MySQL DB서버를 연동했으며, 웹 서버로는 Apache 1.3을 사용했다.

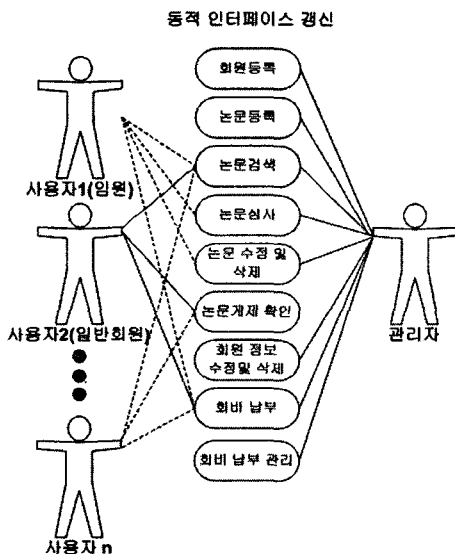


그림 1. 인터페이스 설계

### 2.2 기능

현재 구축되어 있는 기능으로는 등급별 회원관리 기능, 논문열람 및 검색기능, 온라인을 통한 논문 투고기능, 웹 메일러 기능, 로그 기록기능 등이다[2].

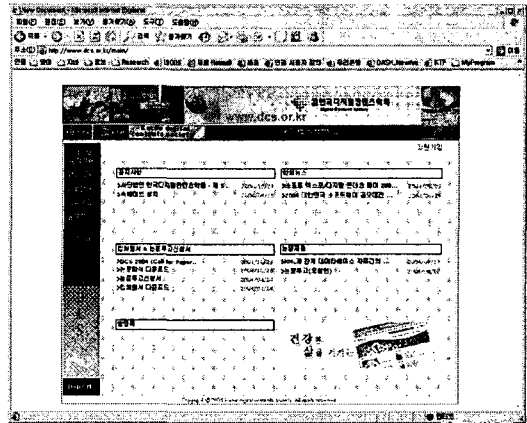


그림 2. 시스템 초기 화면

#### 2.2.1 웹 메일러 기능

웹 메일러 기능은 학회의 공지사항이나 회원들에게 필요한 정보를 원하는 형태로 제공하기 위해 구현됐다. 웹 메일러 기능을 통해서 관리자는 여러 회원들에게 정보를 신속하게 공지할 수 있도록 했으며, 등급별, 관심분야별로 서로 다른 정보를 공지하도록 구현했다.

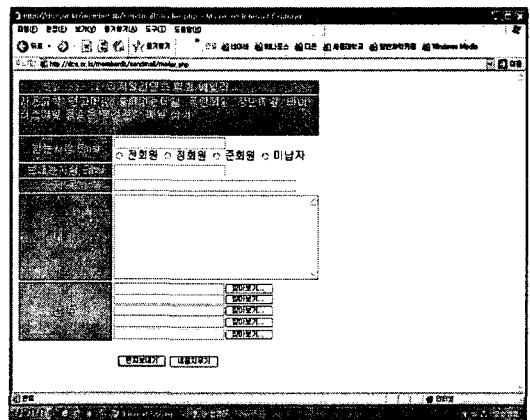


그림 3. 폼 메일러

웹 메일러는 PHP 폼 메일러로 구축했으며 헤더 정보와 메일의 기본 구성 정보는 기본적으로 웹 메일 서버가 자동으로 할당한다. 메일 수신자는 회원 등급, 관심분야 등의 카테고리로 나눠져 있으며 e-mail 정보를 DB로부터 얻어와 메일을 보낼 수 있게 된다. 기본적인 알고리즘은 다음과 같다.

```
#####디비 쿼리 부분#####
if($mailto=="all")$query = "select email from member";
else if($mailto=="regular")
$query = "select email from member where
userlevel='정회원'";
else if($mailto=="associate")
$query = "select email from member where
userlevel='준회원'";
else if($mailto=="nonpay")
$query = "select email from member where fee2004='';
if($query)$result = mysql_query($query);
#####파일첨부 MIME 정보 파트#####
$mailbody.= "--$boundaryWrWn";
$mailbody.="Content-Type:".$userfile_type{$i}.";
name=W".$filename."WrWn";
$mailbody.="Content-Transfer-Encoding: base64WrWn";
$mailbody.="Content-Disposition: attachment;
filename=W".$filename."WrWnWrWn";
$mailbody .= base64_encode($file)."WrWnWrWn";
#####메일 발송 루프#####
while($info = mysql_fetch_array($result,MYSQL_ASSOC)) {
$mail = $info[email];
mail($mail , addslashes($subject) , $mailbody , $header);}

```

2.2.2 회원 관리 기능

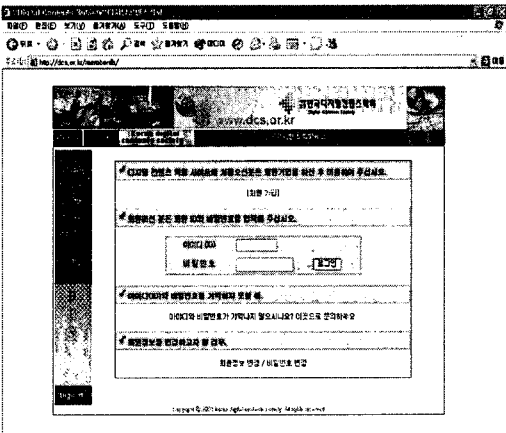


그림 4. 학회 회원 관리 시스템 초기화면

그림 4에 나타난 화면은 학회 사이트의 여러 기능들 중 하나인 학회 회원 데이터베이스 관리의 메인페이지(http://dcs.or.kr/memberdb/)이다. 학회 회원으로 활동을 원하는 임의의 사용자는 [회원 가입] 링크를 클릭하여 학회 회원 데이터베이스에 신상 정보를 입력할 수 있도록 했다. 이미 학회 회원인 사용자는 로그인을 통해 학회 회원 데이터베이스에 연결되어 회원임을 인증받고 학회 사이트의 서비스를 제공받도록 하였다. 아이디나 비밀번호를 잊어버린 회원의 경우 관리자에게 연락할 수 있도록 했으며 웹 메일러를 통해 로그인 정보가 발송 되도록 했다. 또한 링크를 클릭 했을 때 전체

프레임을 refresh 하지 않고 iframe 기법을 이용하여 refresh 속도가 증가되도록 하였다. 좌측 메뉴는 디지털 콘텐츠 학회 사이트 페이지의 공통 링크로서 학회 소개, 임회 안내, 논문지, 학술 대회에 대한 정보를 제공하기 위한 부분이다. 공통 링크는 모든 사용자에게 제공되며 사용자 인증을 통해 등급별 차별화된 링크를 추가적으로 제공할 수 있도록 했다.

다음으로 임의의 사용자가 회원 가입을 요청했을 때 반응하여 나타나는 학회 회원 가입 페이지의 주요 기능을 살펴 볼 것이다.

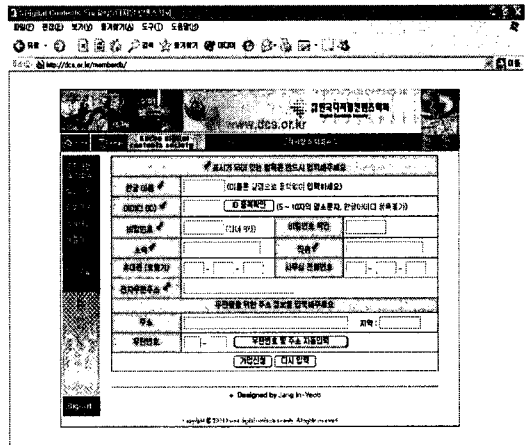


그림 5. 사용자 Profile 입력

그림 5에 나타난 화면은 회원의 기본 정보를 학회 회원 관리 데이터베이스에 실제적인 회원 데이터를 입력하는 것으로 사용자의 정보를 입력하는 것이다. 여기서 입력된 정보들은 필터링을 통해 사용자의 행동 분석에 필요한 정보만을 추출하여 따로 데이터베이스화 한다.

2.2.3 논문 검색 기능

논문 정보 검색 기능은 원하는 논문 정보를 사용자가 빠르게 접근할 수 있도록 구현한 것으로 키워드 기반 검색이다. 사용자가 원하는 논문을 빠르게 검색할 수 있도록 주제, 출판년도, 키워드, 저자명 등 다양한 카테고리를 통해서 검색이 가능하도록 구현되어 있다.

설계 단계에서 향후 구현할 검색 기능으로는 관심 키워드를 등록했을 경우 접속할 때 관련된 논문의 일련번호를 제공해서 요약문만 볼 수 있도록 하는 기능이 있다. 논문 정보 검색 알고리즘은 다음과 같다.

웹에 대한 사용자의 접속과 행동에 대한 기록은 서버에서 USE\_LOG 테이블에 기록하여 관리자 페이지를 통해 리포팅 되도록 했다.

```

if ($_POST[classification]!="all") {
    $query = "SELECT distinct title, s_number,
    author, classification, reference, date,
    subject, publication_date FROM literature
    where title LIKE '%$$_POST[title]%' group by title: ";
} else {
    $query = "SELECT distinct title, s_number,
    author, classification, reference, date,
    subject, publication_date FROM literature
    where title LIKE '%$$_POST[title]%'
    AND classification='$_POST[classification]'
    group by equip_name: ";
}
$result = mysql_query($query); ?>
    
```

2.3 데이터베이스 설계

데이터베이스는 사이트내의 정보를 효율적으로 관리 할 수 있도록 설계 되었으며 크게 회원정보, 기업정보, 분야정보 등을 각각 Member, Department, Topic 이라는 3개의 테이블로 구분했다. 각각의 테이블의 데이터들은 일련번호를 부여하여 구분했다. 특히 회원 정보 테이블은 회원들의 특성을 크게 '일반회원', '임원', '학생회원'으로 분류하여 정보를 분산 시켰다.

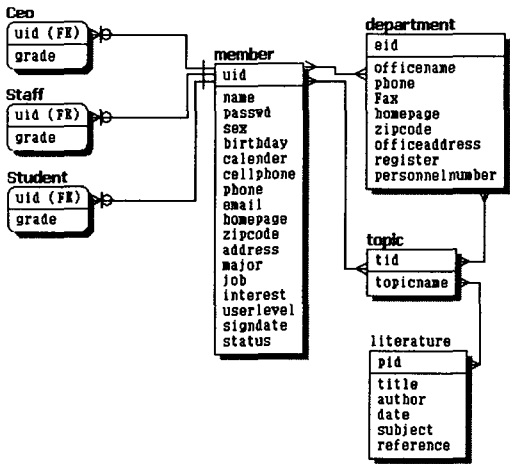


그림 6. E-R Diagram

그림 6 은 설계된 데이터베이스를 Diagram으로 도식화한 것이다. 먼저 회원정보(member)테이블의 스키마는 다음과 같다.

Member(uid, name, passwd, birthyear, birthmonth, birthday, calendar, sex, email, homepage, major, job, addr, zipcode, interest, phone, cellphone, signdate, userlevel)

다음은 Member 테이블의 subclass인 '임원'(CEO), '일반회원'(Generalstaff), '학생회원'(Student)의 테이블 스키마이다. 여기서 '임원'은 회원등급의 값인 '임원'이 아닌 소속(기업)의 임원을 말한다.

- CEO(uid, occupation)
- Staff(uid, occupation)
- Student(uid, occupation)

소속정보(Department)테이블은 다음과 같다

- Department(eid, Officename, Fax, officeaddress, zipcode, register, personnel number, homepage)

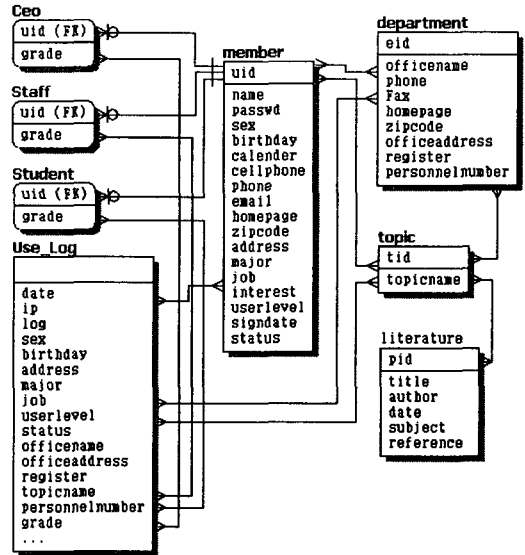


그림 7. 데이터마이닝을 위한 데이터베이스 모델링

그림 7. 은 데이터마이닝을 위한 데이터베이스 모델링 사례이다. 학회에서 보유하고 있는 회원기본정보(Member) 소속정보(Department), 관련분야정보(Topic), 논문(Literature)정보 등의 테이블과 실제 데이터 마이닝에 쓸 data들을 가지고 있는 테이블 USE\_LOG 테이블로 구성되어 있다. USE\_LOG 테이블은 사용자의 웹 사용 정보와 앞서 언급한 필터링을 통해 여러 테이블의 정보 중 데이터마이닝에 유효한 정보들을 구성한 테이블이다.

3. 데이터마이닝을 통한 규칙발견

데이터마이닝은 방대한 데이터를 기반으로 데이터웨어하우스나 데이터마트안에 저장되어 있는 대량의 데이터로부터 쉽게 드러나지 않는 유용한 정보들을 추출하는 과정을 말한다. 대량의 데이터는 분석 없이 그 자체만으로는 의미가 없다. 그 이유는 그 데이터의 용량이 너무 방대하고 복잡해서 보통 사람들이 그 데이터를 이해하기란 매우 힘들기 때문이다. 데이터 분석가들도 그들이 필요한 부분만을 선택하여 심층 분석을 한다. 그러므로 그 이외에 사용하지 않은 데이터들로부터 의미있는 정보나 지식 등을 놓치는 경우가 발생하게 된다. 이러한 지식을 유출해내는 방법은 어떤 특정 기법과 그 기술 자체만을 의미하는 것이 아니고, 비즈니스 문제나 과학용 데이터들을 분석하는 문제 등에서 주어

진 상황을 이해하고 그 특정한 문제를 해결하기 위하여 여러 학문 분야의 방법들을 적용하는 포괄적인 과정을 의미한다. 인터넷과 분산 데이터베이스의 발달로 인한 온라인 데이터들의 급증으로 데이터마이닝의 필요성은 더욱 절실했다. 이미 데이터 마이닝은 비즈니스 분야에서 CRM(Customer Relationship Management:고객 관계 관리)의 핵심적인 기술로 사용되고 있으며 바이오테크놀로지의 발달로 인해 그 분야의 방대한 데이터를 분석하는 기술로 부각되었다[3,4]. 성공적인 데이터 마이닝 시스템 구축을 위해서는 목표에 대한 정확한 이해가 필요하며 구체적인 활용방안이 제시되어야 한다. 또한 레코드의 충실도를 고려해야 하고 도출된 결과에 대한 올바른 해석이 중요하다[5].

3장에서는 지능형 학회 관리 시스템을 통해서 얻어진 데이터를 기반으로 각각의 패턴을 분석해서 회원 관리의 효율성 증진을 위한 데이터마이닝 사례를 제시한다.

**3.1 문제 정의 및 목표 설정**

데이터마이닝을 위해 학회 DB의 레코드를 사용했다. 학회 사이트는(<http://www.dcs.or.kr>) 학회의 수많은 회원들이 방문을 하며 이들 대부분은 디지털 콘텐츠 분야에 대한 연구와 개발에 대해 큰 관심을 가진 사람들이다. 논문 투고 및 심사와 출판 그리고 웹을 통한 배포는 학회에서 매우 신중한 판단이 요구되고 있기 때문에 각각 사용자 특성에 맞는 사용자 위주의 적극적인 관리 시스템의 운영이 필요하다. 일반적으로 학회 회원으로 가입하면 모든 회원들에게 동일한 공지 메일이 발송되며 이는 사용자가 자기에게 필요한 정보만을 선택적으로 수용하기 위해서 수신한 메일들을 걸러내야 하는 수고를 해야만 한다. 또, 사용자마다 동일한 인터페이스를 통해 동일한 정보를 제공함으로써 획일화된 서비스를 제공받게 된다. 앞으로 제시할 데이터마이닝의 목적은 관심분야와 웹 활동 기록을 토대로 한 데이터 마이닝을 수행하여 획일적인 서비스를 지양하고 각 사용자에게 맞는 맞춤형 정보를 제공함으로써 사용자가 원하는 정보를 원하는 형태로 전달받을 수 있도록 하고자 한다.

**3.2 마이닝을 위한 준비 작업(Data Preparation)**

데이터마이닝의 수행 과정은 그 목적에 따라 다양한 방법들이 존재한다. 유용한 패턴과 같은 지식들을 추출하기 전에 데이터를 선택하거나 여과, 통합, 샘플링, Noise 제거, 변형등과 같은 과정을 통해 업무량을 줄이거나 분석의 효과를 높인다.

표본 선정및 추출 작업에서는 앞서 언급한바와 같이 사용자들의 학회 활동 기록과 기본정보를 USE\_LOG 테이블에 기록 했고. USE\_LOG 테이블에서 발생한 결손값은 사용자 Profile 입력시 필터링을 통해 처리했다.

Data Preparation 단계는 데이터마이닝을 위한 준비 단계로 데이터마이닝 엔진을 통해 분석하기 쉽도록 정보를 형식화해서 제공한다. 데이터베이스내의 정보는 contin- uers value 과 symbolic value로 표현되며 데이터 마이닝을 위한 준비 파일인 \*.lea 파일과 \*.ctr파일을 생성함으로써 Data Preparation 결과를 표현한다. \*.ctr파일은 각 사이트의 데이터 feature를 추출한 것으로 마이닝 할 각각의 attribute에 대한 정보와 class에 대한 정보, 그리고 다수의 파라미터에 대한 정보를 제공한다. 파라미터 정보는 데이터 마이닝을 위해 제공되는 정보로 attribute의 범위와 가지치기(pruning)유형 및 단계에 대한 정보, 엔트로피(entropy)정보 등 여러 정보를 제공한다. \*.lea 파일에서는 \*.ctr파일에서 형식화된 attribute의 값을 표현한다.

데이터마이닝에 사용될 데이터의 선택은 앞서 설명한 것처럼 질의를 통해 USE\_LOG 테이블에 기록했다. USE\_LOG테이블에 적재하는 과정에서 발생한 결손 값을 제거하기에 앞서 회원들의 세부 정보와 프로필 정보가 일치하지 않는 레코드를 먼저 선별했다. 이 과정에서 많은 레코드가 줄어들었으며 결손 값을 가지는 원인은 많은 필드가 필수 항목이 아닌 경우 임시로 값을 입력하거나 누락시키는 경우가 많은 것으로 나타났다. 또 데이터베이스의 메타데이터를 기반으로 한 탐색과정에서 나타난 값들을 수치화했다. 수치화한 데이터 타입들은 VARCHAR, CHAR, DATE, ENUM 등이다. 수치화된 데이터들은 가치화 단계에서 원래의 형태로 복구했으며 데이터베이스 내의 모든 데이터들을 임시파일 (\*.lea, \*.ctr)로 생성하여 마이닝을 위한 준비 작업을 끝냈다.

**3.3 모형화(Modeling)**

모형화 단계는 데이터마이닝 과정중 가장 중요한 단계로서 선행 단계에서 선정된 주요 변수를 사용하여 다양한 모형을 적용해 보는 단계이다. 모형화 단계에서는 의사결정나무(Decision Tree) 모형을 적용했다. 의사결정나무는 추출된 모델이나 지식들을 아주 잘 표현할 수 있는 장점이 있기 때문에 데이터 마이닝의 방법론 중에 가장 많이 활용되는 방법중 하나이다. 반면에 데이터가 복잡해 질 때 나무구조가 복잡해지는 점이 단점으로 지적된다.[6]

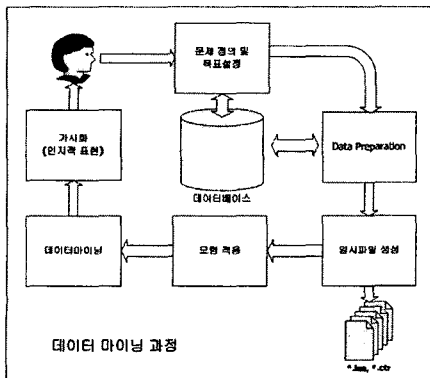


그림 8. 데이터마이닝 수행 과정

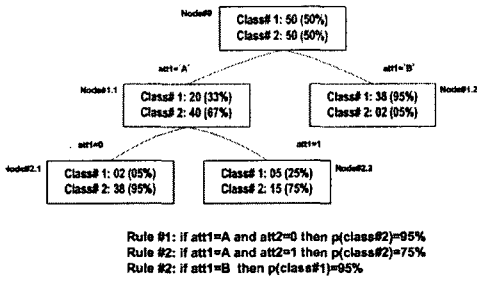


그림 9. 의사결정나무를 통한 분류 예

먼저 회원의 가입정보 및 학회 사이트 활동 로그 정보에 대한 Data Preparation 단계를 거친 후 모형화하여 데이터마이닝을 수행한다. 다음은 모형화를 통해 분류된 여러 군집중 하나의 규칙을 나타낸다.

```

IF(degree='Master')
AND(major=='Software Engineering')
AND(log=='Graphics')
AND(date>=10)
AND(interest=='Augment Reality')
Then->interesting Area 'Computer Graphics'
  
```

분류규칙을 보면 회원의 major(전공)가 Software Engineering이고 Degree(학위)가 석사이며 학회 log(방문기록)에 그래픽관련 정보를 많이 이용하고, date(방문빈도)가 10회 이상이면서 또한 interest(관심분야)가 Augment Reality(증강현실)인 회원이 필요로 하는 정보가 컴퓨터 그래픽이라고 분류된 예를 볼 수 있다. 여기서 주의할 점은 log(방문기록)에는 회원이 이용한 다양한 서비스들이 기록되는데 그 중 가장 많이 이용한 서비스를 추출한 것이고 date(방문빈도)는 다른 조건을 만족하더라도 방문빈도가 적으면 그러한 정보를 필요로 하지 않는 회원일 수 있으므로 매우 중요하다. 이 밖에도 여러 규칙들이 추론됐으며 그중에는 우리가 미처 기대하지 못했던 규칙들도 많이 포함되어 있었다.

3.4 가시화(Visualization)

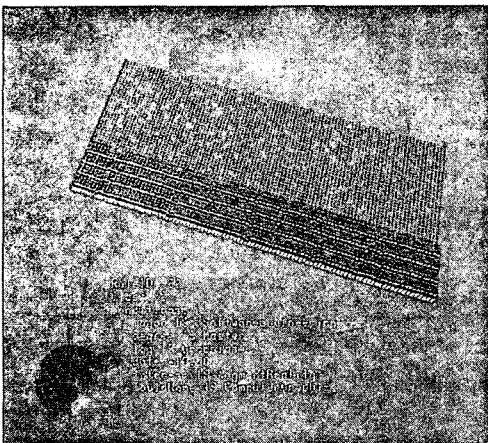


그림 10. InferView를 이용한 가시화(Visualization)

가시화(Visualization)는 데이터마이닝의 수행 결과를 사용자가 보기 편하고 이해하기 쉬운 형태로 제공하기 위해서 각종 차트 형태로 보여주는 것을 말한다.

그림 10 은 모형화 과정에서 얻어진 규칙(rule)들을 사용자에게 효과적으로 전달하기 위해 InferView를 이용해 가시화한 예이다. 그림에서 나타난 규칙은 앞서 언급한 분류조건을 만족하는 규칙들이다. InferView에서는 3차원 구를 통해 분류조건을 만족하는 데이터의 군집을 표현 했으며 구를 클릭 했을때 나타난 3차원 막대그래프는 분류조건을 만족하는 모든 데이터분포를 나타낸다. Visual Pruning(가지치기) 기능을 이용해서 신뢰성을 가지기에는 군집이 미세한 규칙들을 제거할 수 있다[7].

4. 개인화

4.1 개인화의 활용

온라인에서 개인화를 활용하는 것은 실시간으로 개인화 기법을 적용하며 웹사이트의 사용자에게 새로운 것을 소개하고 사용자들의 잠재적 요구를 통해서 웹사이트의 내용을 변화시키는 것을 말한다. 이는 실시간으로 축적된 데이터의 패턴을 분석함으로써 가능하며 이를 위해 실시간으로 데이터 패턴을 분석하는 지능형 개인화 엔진이 필요하다. 개인화 엔진은 인구통계, 구입된 정보, 등록데이터, 다양한 유형의 방문객들이 웹사이트와 상호작용으로 얻은 사용자 데이터 등을 활용한 웹마이닝 기법을 사용한다. 개인화를 통해 제공되는 것을 보면 다음과 같다.

- 사용자의 관심사를 기억해서 제공한다.
- 사용자의 과거의 행동을 기반으로 현재의 행동을 예측해서 그에 맞는 콘텐츠를 보여준다.
- 웹사이트에 처음 방문하는 사용자들을 충성도가 높은 사용자로 바꿔준다.

가장 잘 알려진 예는 Amazon.com으로 웹사이트에서 고객에게 널리 제품을 판매하기 위해 개인화 기법을 사용했다. 개인화는 판매, 마케팅, 고객 서비스가 세 분야에서 실시간으로 고객 상호작용을 지원한다. 따라서 데이터마이닝 어플리케이션과 시스템의 기능적인 연결은 매우 중요하다[8].

4.2 개인화 방안

지능형 학회 관리 시스템의 개인화 방안은 사용자별로 서로 다른 인터페이스를 제공하며 이를 기반으로 한 추천 시스템(Recommendation System)을 구현하기 위한 설계를 제시한다. 예를 들면, 기존 회원이 학회 회비를 미납한 경우 웹 메일링 서비스를 통해 일괄적으로 회비 납부 공지를 보내고, 웹에 로그인시 회비납부 요청 공지를 띄운다. 모든 사용자 정보는 데이터베이스에 기록되기 때문에 회원 가입시 필터링을 통해 profiling 된 정보와 USE\_LOG 테이블에 적재된 사용자의 웹 활

동기록 그리고 카테고리별로 분류된 사용자 정보를 이용하여 간단히 구현할 수 있다. 사용자의 관심도에 따라 사용자가 실제로 관심을 가지거나 잠재적으로 관심을 가질 것으로 예측되는 분야의 정보를 추천하는 시스템을 구현하기 위해서는 앞서 분석된 내용을 활용해야 한다. 즉, 사용자 특성에 따라 분류된 관심분야와 데이터베이스에 기록된 실제 학회 활동 기록을 토대로 사용자 그룹을 정의한다. 정의된 사용자 그룹별로 관심 분야 정보를 추천하도록 구현한다. 또 다른 방법은 사용자의 클릭을 분석해 사용자별 관심 분야 정보를 알아낸다. 개인화를 웹 사이트에 통합하기 위해 상위레벨에서 필요한 것들을 보면 다음과 같다.

- 규칙 데이터베이스 : 여기에는 어떤 조건에 어떤 일을 하도록 하는 정보 뿐 아니라 각종 규칙들도 넣는다.
- 이미 정의된 사용자 그룹 : 몇 가지 영역으로 분류한 사용자 그룹을 저장해 둔다.
- 클라이언트 커뮤니케이터(Client Communicator) : 사용자가 하는 일에 따라 적절한 말을 해 준다.
- 컴포넌트 아키텍처(Component Architecture) : 서버 프로세싱을 하다 보면 위의 세 가지 방법을 구현하기 위해 컴포넌트 아키텍처가 필요하다.[9]

그림 11 은 개인화 과정과 개인화를 위한 시스템 구조를 나타낸 것이다.

사용자 로그인시 사용자 특성에 따라 메뉴의 구성과 링크를 바꿔주도록 하기 위해서는 홈페이지의 메뉴구조, 디자인요소, 실제 제공되는 콘텐츠 등을 분리해서 관리해야 한다. 즉, 서로 다른 디자인 템플릿을 적용해 동일한 콘텐츠를 여러 메뉴에 제공할 수 있어야 하고 상이한 채널을 통해 출판할 수 있도록 해야 하며[10] 사용성(Usability)을 고려한 인터페이스 관리가 필요하다.

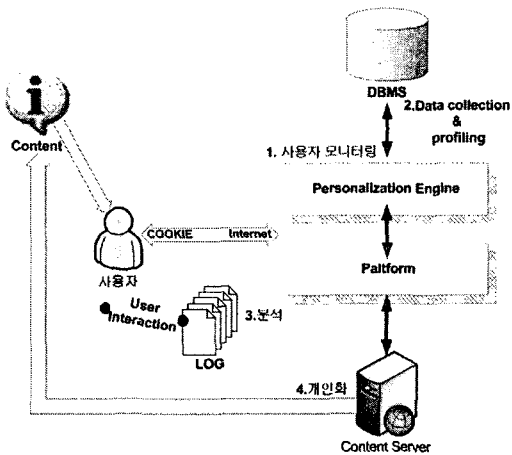


그림 11. 개인화 과정 및 시스템 구조

## 5. 결론 및 향후 방향

본 논문에서 제시한 지능형 학회 관리 시스템의 설계 및 구현 사례는 시스템을 통해서 얻은 학회 웹사이트 사용 기록, 학회 DB에서 얻은 회원의 세부 정보를 분석했다. 추출된 데이터를 기반으로 데이터마이닝 기법을 적용해 몇 가지 중요한 패턴을 발견했으며 이를 통한 개인화 방안, 즉 각 회원의 필요 정보를 제시했다. 분석된 정보들은 학회 운영에 도움이 될 만한 정보들으로써 학회 사이트와 회원간에 실질적인 상호작용을 위해 유용한 작업이었다고 생각한다. 그러나 모형화 과정에서 의사결정나무만을 사용한 점에서는 보다 정확하고 신뢰도 높은 분석 결과를 위해, 데이터마이닝의 목적이거나 데이터의 특징에 맞는 기법들이 선택되어야 하므로 신경망, 클러스터링 등의 여러 기법들을 적용하는 노력이 필요하다. 본 논문에서 데이터마이닝 과정을 통해 추천 시스템 구축을 위한 관심분야 분석 작업으로 개인화 시스템 설계를 제시했다. 이를 토대로 실시간 데이터마이닝을 이용한 개인화 솔루션을 구현하기 위한 연구와 개발이 필요하다.

## 참 고 문 헌

- [1] Lingras, P, "Rough set clustering for Web mining", FUZZ-IEEE Proceedings of the 2002 IEEE International Conference, Vol. 2, pp. 1039 - 1044, 2002
- [2] <http://www.dcs.or.kr>
- [3] 백성욱, "데이터마이닝 분야의 연구, 개발 및 활용", KOSEN([www.kosen21.org](http://www.kosen21.org)), 2001
- [4] 백성욱, "데이터마이닝의 활용에 관한 기술조사", KOSEN([www.kosen21.org](http://www.kosen21.org)), 2000
- [5] Alex Berson and Stephen Smith, "Building Data Mining Application for CRM", Kert Therling, 1999
- [6] Fadila B, Jerome D, "Decision Tree Modeling with Relational Views", LNCS, Vol. 2336, pp. 27-29, 2002
- [7] Bala J, Baik S, Gutta S., Hadjarian A., Mannucci M., and Pachowicz P, "InferView: An Integrated System For Knowledge Acquisition And Visualization", proceedings of the Federal Data Mining Symposium & Exposition 99, McLean, Virginia, 1999
- [8] "Technology Forecast: 2002-2004", Vol. 1: Navigating The Future of Software
- [9] Magdalini Eirinaki, Michalis Vazirgiannis, "Web mining for web personalization" ACM Transactions on Internet Technology (TOIT), Vol. 3, Issue. 1, 2003
- [10] Jiawei Han, Chang, K.C.-C, "Data Mining for Web Intelligence", Computer, Vol. 35, Issue. 11, pp. 64 -70. 2002



### 조 영 기

2004 세종대학교 컴퓨터공학과  
(공학사)

현재 세종대학교 디지털컨텐츠학과  
석사과정

관심분야 : 컴퓨터 비전, 콘텐츠관리시스템,  
데이터마이닝



### 백 성 옥

1987년 서울대학교 계산통계학과  
(학사)

1992년 미국 Northern Illinois  
University (석사)

1999년 미국 George Mason  
University (박사)

1997~2002년 미국 워싱턴 DC 소재 Datamat  
Systems Research Inc.

2002~현재 세종대학교 컴퓨터공학부 디지털컨텐츠학  
과 교수 Email: sbaik@sejong.ac.kr



### 방 기 천

1981년 서울대학교 전자공학과(학사)

1988년 성균관대학교 정보처리학과  
(석사)

1996년 성균관대학교  
전산통계학전공(박사)

1984년~1995년 MBC 기술연구소

1995년~현재 남서울대학교 멀티미디어학과 교수

관심분야 : 멀티미디어컨텐츠, 멀티미디어 응용,  
인터넷 방송 등